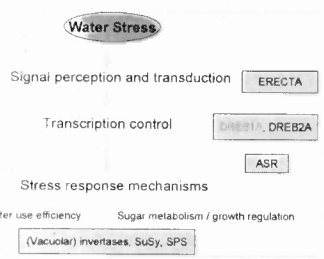
Some Data on Allele Diversity at Orthologous Candidate Genes in GCP Crops (ADOC project, GCP2006-02)

Dominique This[1], Brigitte Courtois[1], Romain Philippe[1], Pierre Mournet[1], Claire Billot[1], Jean-Christophe Glaszmann[1], Roland Schaffalitzky[2], Reinhardt Simon[2], Percy Rojas[2], Merideth Bonierbale[2], Rajeev Varshney[3], C. Tom Hash[3], Hari Upadhyaya[3], Spurthi Nayak[3], Dominique Brunel[4], Redouane El Malki[4], Marie Christine Le Paslier[4], Kenneth McNally[5], Michael Baum[6], Wafaa Choumane[6], Maria Von Korff[6], Matthew Blair[7], Martin Fregene[7]

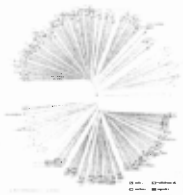
[1] CIRAD - UMR DAP 1098, CIRAD TA A-96/03, Av. Agropolis, 34398 Montpellier cedex 5, France
[2] CIP P.O. Box 1558 La Molina, Lima 12, Lima, Peru
[3] ICRISAT Patancheru, Andhra Pradesh 502 324, India
[4] INRA, UR 1279 Etude du Polymorphisme des Génomes Végétaux, CEA Institut de Génomique/ Centre National de Génotypage, 2, rue Gaston Crémieux, CP5724, Evry F-91057, France
[5] IRRI DAPO Box 7777 Manila 1301, Philippines
[6] ICARDA P.O. Box 5466 Aleppo, Syrian Arab Republic
[7] CIAT km 17 Recta Cali-Palmira, AA6713, Cali, Colombia



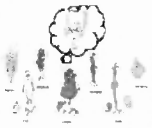
Six gene families were selected as the initial subset of target genes. They act at different levels of the drought stress response.

	ASR	VAI	ER/ERL	DREB2A /2B	SuSy	SPS	totalcrop	Length	Depth	Total bp
rice	5	1	3	1	7	1	18	2/261	131to280	4307772
barley	3	3	2	1	1	1	11	8/414	143to274	1830037
sorghum	5	1	3	0	4	1	14	12/065	127to347	3422803
cassava	1	0	3	0	4	1	9	6/531	153to281	1405923
potato	3	0	1	0	2	1	7	5/744	64to289	1004494
chickpea	1	0	1	1	0	1	4	4/09	78to235	622990
bean	2	0	1	2	2	0	7	6/631	60to241	1182143
totalgene	20	5	14	5	20	6	70	65/585	60to347	10Mbp

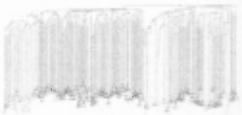
Around 10 Mbp sequences were generated and compared within and across species. Some new orthologs were identified by using degenerate primers and reconciliated taxonomic trees (Nayak et al, 2009), and allelic series derived by specific PCR



Population structure of the rice reference collection – K McNally (IRRI), B Courtois, R Philippe (CIRAD)

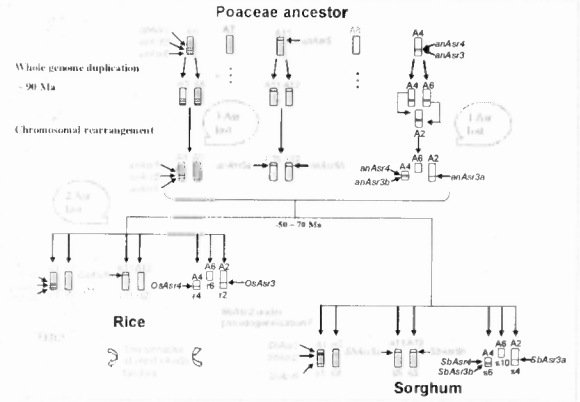


The seven crops studied in this project are barley (Hv), sorghum (Sb), rice (Os), potato (Stbr), cassava (Me), bean (Pv) and chickpea (Ca).
Faced to drought problems, they are supposed to have learned how to use more or less adapted tools, its adapt gene structure and expression
(drawing from Camille This, 2006)



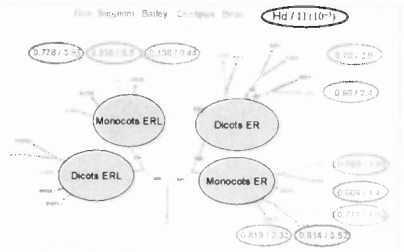
Population structure of CIP potato CGS.
From R. Herrera and J. Nunez CIP (The potato reference collection included polyploids)

Allelic sequences were generated in reference collections of seven crop species considered as GCP priorities (around 300 accessions for each crop)
Faced to drought-prone environments in some area, different selection strategies may have shaped the diversity patterns of our target genes... Population structure also reflects domestication and selection's history, to be taken into account.

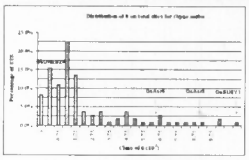


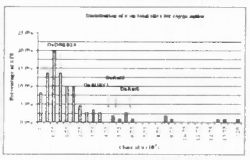
This evolution model for Asr gene family is based on Salse et al (2008) model of poaceae genome evolution, simplified in order to represent only chromosomal regions harbouring Asr genes, in a comparative analysis between rice and sorghum. Orthology relationship between rice and sorghum was inferred from phylogenetic and microsynteny analysis (Philippe et al, in preparation).

Comparison between whole sequenced genomes allow some comparison between orthologs and paralogs within a gene family - Here the Asr (ABA- stress-ripening) family. Some hypothesis can be proposed, following Salse et al. (2008) model for poaceae genome evolution, to infer a minimal gene number in the poaceae ancestor (6 for Asr). Several Asr were lost following whole genome duplication and in the rice lineage. A translocation event disrupted the microsyntheny between rice and sorghum around Asr1-Asr2 tandem. NB: in agreement with data on *Brachypodium* and maize (not shown here)



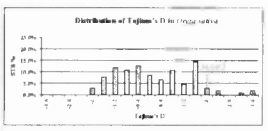
Phylogenetic tree of the ERECTA and ERECTA-like gene family (established in collaboration with J. Masle, ANU), showing diversity levels (Hd: haplotype diversity and π: nucleotide diversity) for several orthologs or paralogs. SbERL diversity is much lower than orthologs (OsERL and HvERL) and paralogs (SbER1 and SbER2)

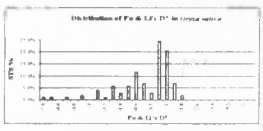


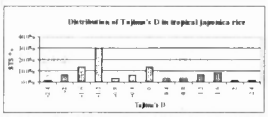


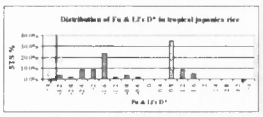
Candidate genes outliers in rice, based on data at the genome level from Caicedo et al (2007) (111 randomly chosen STS). Two diversity indices (π and H) are shown.

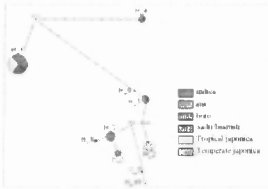
Diversity level of our target genes differed clearly between the different genes and gene family members within a crop, and between orthologs when comparing several crops (here for ERECTA and ERECTA-like genes).
A few outliers were identified, compared to average diversity data. DREB2A presented a very low diversity in all species, while barley's vacuolar invertase-related sequences and potato's StbrSUSY1 presented a very high diversity within the reference collections analyzed.



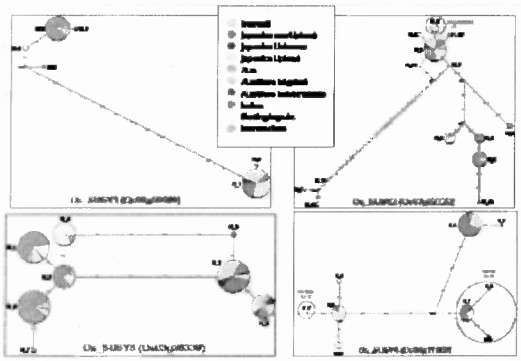








Tests for selection were performed to estimate whether the considered genes followed the model of neutral evolution. D (Tajima, 1959) and D* (Fu and Li's, 1993) are shown here for OsAsr3, and compared to randomly chosen data from Caicedo et al (2007), either for *Oryza sativa* globally (up), or for the temperate japonica subgroup (down).
The haplotype network of OsAsr3 in cultivated rice shows a much higher diversity in indica than japonica, and a very distinct haplotype (H4) for a few accessions. (Philippe et al, submitted)



Haplotype networks are constructed for 4 rice sucrose synthase genes using NETWORK v4.5. Each node represents an haplotype and segments correspond to mutations (SNP or Indel). The different colors represent the varietal groups.

Haplotype networks are sometimes influenced by population structure
(An example here for OsSUSY1 and OsSUSY4, where a clear distinction can be made between japonica and indica accessions, except putative introgressions),
but the diversity pattern is less clear for some other genes like OsSUSY3.
A few non-synonymous changes are highlighted here for OsSUSY2 and OsSUSY4.

When computing indices for selective pressure on our data, a few candidate genes were highlighted (here OsAsr3, suggesting a balanced selection at the species level, and a directional selection within the tropical japonica group). The haplotype network of this gene may provide some clues to explain this feature.
Other candidate outliers include HvAsr5, OsSUSY1, MeSUSY3 and PvERECTA (Positive D), StbrAsr1 and SbAsr1 (Negative D)

This study highlights the complexity of evolutionary and diversity patterns of candidate genes within gene families.
Genome evolution as well as population structure and some selection pressure have shaped gene's diversity.
Functional inference based solely on orthology relationship should be therefore considered with caution.
Functional analysis of haplotypes identified in this research work and eco-geographical data are now required in order to define potential candidate genes and favourable alleles for drought tolerance in GCP crops.